

A non-local Random Walk on the Hypercube

Evita Nestoridi

July 20, 2015

Contents

1	Introduction	2
2	The history of the Ehrenfest's urn model	4
3	Coupling Argument	5
4	Proof of Theorem 1	6
4.1	The upper bound for $k = \frac{n}{2}$	6
4.2	Upper Bound	11
4.3	Lower Bound	15
5	Fourier Transform Arguments	18
6	The case $k=n/2$	19
7	Similar Random Walks	20
8	A random walk on $(\mathbb{Z}/m\mathbb{Z})^n$	20
9	Comparison Theory Application	22
10	Acknowledgements	23

1 Introduction

In the course of studying how to shuffle big decks of cards, G. White and I had to answer the following question: how long does it take to mix n red balls and n black balls, half of which are contained in one urn and the rest in another, if at each step we pick k balls from each urn and exchange them. The answer to that question can be found in [11]. To study the above model, it was natural to ask a similar question about Ehrenfest's urn model. More precisely, consider two urns. Initially urn one contains zero balls and urn two contains n balls. At each step, pick k total balls at random and move each of them to the opposite urn.

The above Markov chain can be also viewed as a random walk on $(\mathbb{Z}/2\mathbb{Z})^n$ where at each step we flip k random coordinates (for some fixed k). For the walk to be transitive, k needs to be odd, and to avoid parity problems, it is simplest to consider the lazy version of this walk. In other words, at each step, do nothing with probability $1/2$ and with probability $1/2$ choose a random set of k coordinates and flip them. The main question is to find the mixing time of this walk for the total variation distance.

This non-local walk implies a big change at each step. So far people have mostly studied local models; in the case of the hypercube for example, the most famous model is the one that considers picking one coordinate at random and flipping it. But of course in that case the outcome after one step is not very different from the initial configuration, which is why mixing is slower. Of course, really big changes (e.g. $k = n$) make the mixing faster. A first heuristic explained below is that flipping k coordinates at each step should be roughly the same as moving one each time and repeating k times.

There is a second reason why this particular random walk is interesting. There are two different approaches to finding the mixing time of this walk. The first approach is developed in Section 5. It involves finding the eigenvalues of the walk using representation theory and using the Fourier transform to give bounds on the l^2 norm of the difference $P^{*l} - U$. For the case of $k = 1$, this technique works nicely and gives a sharp upper bound on the mixing time. However, for $k = \frac{n}{2}$, it turns out that the bound obtained via the l^2 norm does not give a sharp upper bound on the mixing time, which is defined in terms of the total variation distance (l^1 norm).

A second argument via coupling is introduced in Section 3. It provides a solution to the general case and makes the difference between the l^2 norm and total variation distance clear. This coupling argument is a generalization

of one used by D. Aldous [2] for the case $k = 1$. See [1] for more results of Aldous on the hypercube. The lower bound uses the first two eigenvectors and eigenvalues of the random walk and the second moment technique. This method was firstly introduced by P. Diaconis and M. Shashahani in [7]. In their paper, they managed to prove a lower bound for the case $k = 1$ that matched the Aldous' upper bound, proving in this way the existence of a cut-off at $\frac{1}{4}(n + 1) \log n$. Another way to find a lower bound was proved by L. Saloff-Coste in [13] using Wilson's lemma. In [5], Diaconis, Graham and Morrisson use Fourier Analysis directly to derive the exact behavior of the error for the nearest neighbor random walk.

The results of this section are the following:

Theorem 1. *For the lazy walk changing $k \leq n/2$ coordinates on the n -dimensional hypercube the following hold:*

$$1. \text{ For } l = \left(\frac{8n}{k} \log n + \frac{3n}{2k} + \frac{\sqrt{2n}}{(\sqrt{2}-1)^k} + 2 \right) + c\sqrt{\frac{n}{k} \log n},$$

$$\|P^{*l} - U\|_{T.V.} \leq \frac{1}{c^2}$$

where $c > 0$.

$$2. \text{ For } l = \frac{n}{2k} \log n - c\frac{n}{k}, \text{ where } 0 < c \leq \frac{1}{4} \log n,$$

$$\|P^{*l} - U\|_{T.V.} \geq 1 - \frac{B}{e^{4c}}$$

for a uniformly bounded constant $B > 0$.

Remark 2. *It is easy to see that the mixing time for the k model and the $n - k$ model will be the same, therefore we will focus on the case $k \leq n/2$.*

Here are a few computations:

Section 7 contains the analysis for l^2 -mixing time of the random walk on $(\mathbb{Z}/m\mathbb{Z})^n$ generated by the measure

$$Q(a_{i_1}e_{i_1} + a_{i_2}e_{i_2} + \dots a_{i_k}e_{i_k}) = \frac{1}{\binom{n}{k}m^k}$$

where $a_{i_j} \in \mathbb{Z}/m\mathbb{Z}$ and $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$.

The main result is:

Table 1: Examples for the k -walk on the hypercube

n	k	upper bound for the mixing time
54	27	19
54	3	576
418	209	26
418	7	2899
550	275	27
550	25	1,112

Theorem 3. *For the walk generated by Q , if $l = \frac{n+1}{2k} \log(mn) + \frac{c(n+1)}{2k}$ then*

$$4\|Q^{*l} - U\|_{T.V.}^2 \leq e^{-c}$$

It is known that the l_1 -mixing time is faster than that (using the fact that the first time that we have touched all coordinates is a strong stationary time) but the above result holds for the l_2 norm, which allows us to use comparison theory to provide bounds for the l_2 -mixing times of the walk on $(\mathbb{Z}/m\mathbb{Z})^n$ generated by

$$P_2(\pm e_i) = \frac{1}{4n}, P(id) = \frac{1}{2}$$

More precisely, we prove a bound of the form $m^2(\frac{n+1}{2} \log(mn) + \frac{c(n+1)}{2})$ for the mixing time of the last random walk. The details are included in Section 9. The analysis of l^2 norm of the last walk has already been done by Diaconis and Saloff-Coste [6], where they proved an upper bound of order $m^2 n \log n$ and then Saloff-Coste proved the cut-off [12].

2 The history of the Ehrenfest's urn model

The Ehrenfest's urn model was introduced by Tatjana and Paul Ehrenfest [8] to study the second law of thermodynamics. This is a model for n particles distributed in two containers and each particle changes container independently from the others at rate λ . This process is repeated several times and the question is to find the limiting distribution of the process. M. Kac [9] approached this problem by finding the eigenvectors and eigenvalues of the transition matrix. He also proved that if the initial system state is not at equilibrium then the entropy is increasing.

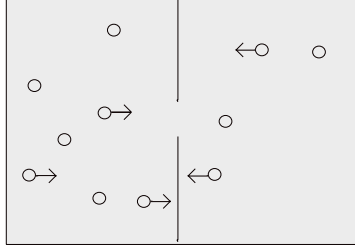


Figure 1: Eleven particles in two containers, five of each changing containers

This problem can also be viewed as a random walk on $(\mathbb{Z}/2\mathbb{Z})^n$ where the ones in a binary vector represent the number of particles in the right hand container. Flipping one (or k) coordinates of the binary vector corresponds to moving one (or k) particles to the other container. But now the Markov Chain problem can be studied through a random walk on an abelian group, where representation theory is quite simple to use. As Persi Diaconis writes in Chapter 3 of his book [4], Kac posed the question: When can a Markov chain be lifted to a random walk on a group?

3 Coupling Argument

Consider the walk on the hypercube $(\mathbb{Z}/2\mathbb{Z})^n$:

$$P(g) = \begin{cases} \frac{1}{2}, & \text{if } g = id \\ \frac{1}{2\binom{n}{k}}, & \text{if } g \in (\mathbb{Z}/2\mathbb{Z})^n \text{ has } k \text{ ones and } N - k \text{ zeros} \end{cases}$$

Here is the coupling argument which will provide an upper bound for the mixing time for $k \leq \frac{n}{2}$:

Start with two different copies of the Markov Chain. At time t denote the state of each as X_1^t and X_2^t . X_1 will start at the identity while X_2 will start at a random configuration. At time t , let

$$y(t) = ||X_1^t - X_2^t||_1 = \sum_i |X_1^t(i) - X_2^t(i)| \quad (1)$$

denote the l^1 distance between the the two configurations. Then consider the following cases:

1. If $y(t)$ is odd then take one independent step on each chain according to the probability measure P .

2. If $y(t)$ is even then with probability $\frac{1}{2}$ stay fixed in both chains. With probability $\frac{1}{2\binom{n}{k}}$ choose k coordinates and change X_1^t completely. In terms of X_2^t , look at the k coordinates that you picked.

Definition 4. Denote by $a(t)$ the number of the mismatching coordinates among the k ones selected.

If $a(t) > \frac{y(t)}{2}$ then change X_2^t completely on the k coordinates picked. Otherwise, if $a(t) \leq \frac{y(t)}{2}$ at first change X_2^t at the coordinates that the two chains match among the k ones picked. Then for every mismatching coordinate among the k ones picked, find the next mismatched (and not found) mismatched coordinate out of the k ones picked, moving cyclically.

Then the following lemma, which can be found in Chapter 4 of [4], says how the above coupling can be used to get an upper bound for the total variation distance:

Lemma 5. Let T to be the first time the two chains match.

$$\|P_{C_0}^{*l} - U\|_{T.V.} \leq P(T > l).$$

The above lemma and Chebychev's Inequality will be the main tools to prove Theorem 2.

Remark 6. For the random walk on a group, both the l^1 and l^2 norms are independent of the starting state. The first chain could start at any fixed configuration. This is why Theorem 1 is stated for any starting configuration.

4 Proof of Theorem 1

4.1 The upper bound for $k = \frac{n}{2}$

Let's take a look at the case of $k = \frac{n}{2}$, for n even: it gives insight for how to prove the general case. Also, the Fourier Transform argument in Section 5 gives a worse upper bound.

Theorem 7. If $l = (8 \log(\frac{n}{2}) + c \sqrt{\log \frac{n}{2}})$ for $c > 0$ then

$$\|P^{*l} - U\|_{T.V.} \leq \frac{1}{c^2}$$

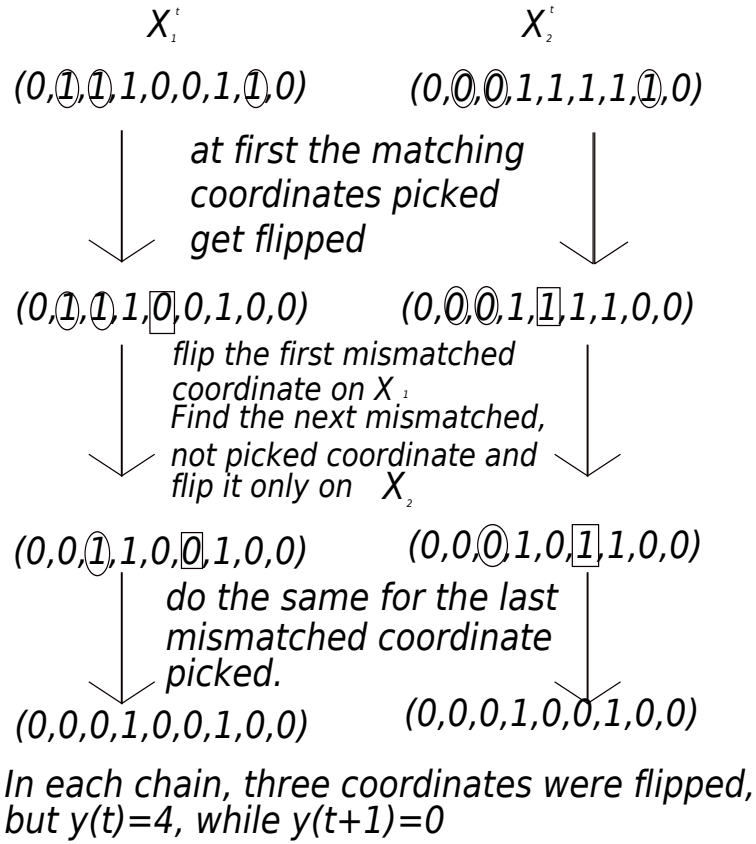


Figure 2: This picture gives an example of how the coupling would work.

The following Lemma will help proving theorem 7: Let X_1^t and X_2^t denote the configuration of the first copy and the second copy of the same Markov Chain respectively. Also, $X_1^0 = id$ while X_2^0 is random. Denote by y_t the number of coordinates that $X_1^t = id$ and X_2^t differ at on time t . Let a_t count how many differing coordinates are picked at step t after starting running the coupling process.

Lemma 8. *With the notation above,*

1. *If $y_t \geq \frac{n}{2}$ then $P\left(y_t - \frac{n}{2} \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{1}{4}$.*
2. *If $y_t \leq \frac{n}{2}$ then $P\left(\frac{y_t}{4} \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{1}{4}$*

Proof.

1. At first notice that if $y_t - \frac{n}{2} \leq i \leq \frac{y_t}{2}$ then

$$\binom{y_t}{i} \binom{n-y_t}{\frac{n}{2}-i} = \binom{y_t}{y_t-i} \binom{n-y_t}{\frac{n}{2}-y_t+i} \quad (2)$$

Therefore using the facts that

$$\frac{1}{2} = P\left(y_t - \frac{n}{2} \leq a_t \leq \frac{n}{2}\right) = P\left(y_t - \frac{n}{2} \leq a_t \leq \frac{y_t}{2}\right) + P\left(\frac{y_t}{2} < a_t \leq \frac{n}{2}\right) \quad (3)$$

and that

$$P(a_t = i) = \frac{\binom{y_t}{i} \binom{n-y_t}{\frac{n}{2}-i}}{2 \binom{n}{\frac{n}{2}}} \quad (4)$$

equations 2,3 and 4 give exactly that

$$P\left(y_t - \frac{n}{2} \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{1}{4}$$

2. For the case where $y_t \leq \frac{n}{2}$ the goal is at first to prove that

$$\frac{1}{4} \leq P\left(1 \leq a_t \leq \frac{y_t}{2}\right) \quad (5)$$

and then that

$$P\left(\frac{y_t}{4} \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{1}{2} P\left(1 \leq a_t \leq \frac{y_t}{2}\right) \quad (6)$$

To prove equation (5) just notice that equation (2) is still valid for $0 \leq i \leq \frac{y}{2}$. Therefore

$$\frac{1}{2} = 2P\left(1 \leq a_t \leq \frac{y_t}{2} - 1\right) + 2P(a_t = 0) + P\left(a_t = \frac{y}{2}\right)$$

Then notice that $2P(a_t = 0) = 2\binom{n-y}{\frac{n}{2}} \leq \binom{y}{\frac{y}{2}}\binom{n-y}{\frac{n}{2}-\frac{y}{2}} = P\left(a_t = \frac{y}{2}\right)$ and therefore equation (5) holds.

To prove equation (6) it suffices to prove that

$$P\left(\frac{y}{4} \leq a_t \leq \frac{y_t}{2}\right) \geq P\left(1 \leq a_t \leq \frac{y_t}{4}\right)$$

which is equivalent to showing that the following inequality holds:

$$\begin{aligned} & \left(\frac{3y}{4}+i\right)\left(\frac{3y}{4}+i-1\right)\dots\left(\frac{3y}{4}-i+1\right)\left(\frac{n}{2}-\frac{y}{4}+i\right)\left(\frac{n}{2}-\frac{y}{4}+i-1\right)\dots\left(\frac{n}{2}-\frac{y}{4}-i+1\right) \\ & \geq \left(\frac{y}{4}+i\right)\left(\frac{y}{4}+i-1\right)\dots\left(\frac{y}{4}-i+1\right)\left(\frac{n}{2}-\frac{3y}{4}+i\right)\left(\frac{n}{2}-\frac{3y}{4}+i-1\right)\dots\left(\frac{n}{2}-\frac{3y}{4}-i+1\right) \end{aligned}$$

This is obviously true since $\frac{3y}{4} + i - b \geq \frac{y}{4} + i - b$ and $\frac{n}{2} - \frac{y}{4} + i - b \geq \frac{n}{2} - \frac{3y}{4} + i - b$.

□

The above lemma is the main tool to the proof of Theorem 7.

Proof. At first, if $y(0)$ is odd then wait until the coupling suggests staying fixed in one of the chains while making moves on the other chain. The expected time for this to happen is 2 since this time follows the geometric distribution with probability of success $\frac{1}{2}$.

To prove the theorem, consider the following two cases.

1. If $y_t \geq \frac{n}{2}$ then consider a much slower process where if at time t you pick less than $y_t - \frac{n}{2}$ mismatching coordinates you do nothing, while if you picked more than $y_t - \frac{n}{2}$ then you act as if you picked only $y_t - \frac{n}{2}$ of them. Now, consider W to be the first time that $a_t = y_t - \frac{n}{2}$ mismatched coordinates are picked then part 1. of Lemma 8 gives that

$$E(W) \leq 4$$

where 4 is the expectation of the geometric with probability of success equal to $\frac{1}{4}$.

This means that after an average of 4 steps the number of mismatched coordinates is $y_{t+2} = y_t - 2(y_t - \frac{n}{2}) = n - y_t$ which in this case would be at most $\frac{n}{2}$.

2. If $y_t \leq \frac{n}{2}$ then again consider a much slower process where if at time t the number of mismatched coordinates picked is a_t and $\frac{y_t}{4} \leq a_t \leq \frac{y_t}{2}$ then act as if you picked $\frac{y_t}{4}$ mismatched coordinates, otherwise do nothing. Let B_i be the i^{th} time that $\frac{1}{4}$ of the mismatched coordinates is picked and i_0 be the time we picked $\frac{1}{4}$ of the mismatched coordinates and we ended up with only 2 mismatched coordinates left. Then $1 \leq i_0 \leq \log \frac{n}{2}$ and if $B_{i_0} = B_1 + \sum_i (B_{i+1} - B_i)$ is the total steps needed to end up with only 2 mismatched coordinates left then

$$E(B_{i_0}) \leq 8 \log \frac{n}{2}$$

Finally, it is important to estimate the probability of picking one mismatched coordinate when there are only 2 mismatched coordinates in order to finish the proof

$$P(a_t = 1) = \frac{1}{2} \frac{\binom{2}{1} \binom{\frac{n}{2}-1}{\frac{n}{2}-1}}{\binom{n}{\frac{n}{2}}} = \frac{n}{4(n-1)} \geq \frac{1}{4}$$

so again if R is the first time one of the two mismatched coordinates is picked then

$$E(R) \leq 4$$

Summing up, if T is the coupling time then

$$E(T) \leq E(W) + E(B_{i_0}) + E(R) \leq 8 \log \left(\frac{n}{2} \right) + 10$$

and

$$Var(T) \leq A \log \frac{n}{2}$$

where A is a constant. Then Chebychev's inequality implies the rest. \square

4.2 Upper Bound

The following lemma is the key to the proof of Theorem 1. The lemma mainly bounds from below the probability that a sufficient number of mismatched coordinates are picked.

Lemma 9. *With $a(t)$ defined in Definition 4 and $y_t := y(t)$ which is given by equation 1,*

1. $P(a_t = i)$ is increasing as a function of i when $i \leq \frac{y_t k - n + y_t + k}{n+1}$ and decreasing when $i \geq \frac{y_t k - n + y_t + k}{n+1}$.
2. If $y_t \geq k$ and $\frac{y_t k}{n} \geq 2$ then $P\left(\frac{y_t k}{2n} \leq a_t \leq \min\{\frac{y_t}{2}, k\}\right) \geq \frac{1}{8}$.
3. If $y_t \geq k$ and $1 \leq \frac{y_t k}{n} < 2$ then $P\left(1 \leq a_t \leq \min\{\frac{y_t}{2}, k\}\right) \geq \frac{1}{6}$.
4. If $y_t \geq k$ and $\frac{\log 2}{2} \leq \frac{y_t k}{n} < 1$ then $P\left(1 \leq a_t \leq \min\{k, \frac{y_t}{2}\}\right) \geq \frac{\sqrt{2}-1}{4\sqrt{2}}$.
5. If $y_t \geq k$ and $\frac{y_t k}{n} < \frac{\log 2}{2}$ then $P\left(1 \leq a_t \leq \min\{\frac{y_t}{2}, k\}\right) \geq \frac{y_t k}{8n}$.
6. If $y_t < k$ and $\frac{y_t k}{n} \geq 2$ then $P\left(\frac{y_t k}{2n} \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{1}{8}$.
7. If $y_t < k$ and $1 \leq \frac{y_t k}{n} < 2$ then $P\left(1 \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{1}{6}$.
8. If $y_t < k$ and $\frac{\log 2}{2} \leq \frac{y_t k}{n} < 1$ then $P\left(1 \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{\sqrt{2}-1}{4\sqrt{2}}$.
9. If $y_t < k$ and $\frac{y_t k}{n} \leq \frac{\log 2}{2}$ then $P\left(1 \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{y_t k}{8n}$.

Proof. The proof is quite technical and goes as following:

1. Let $f(i) = P(a_t = i)$, where $a(t)$ is the number of the mismatching coordinates among the k ones selected. Then

$$f(i) = \frac{\binom{y_t}{i} \binom{n-y_t}{k-i}}{\binom{n}{k}}$$

and $\frac{f(i)}{f(i+1)} \leq 1$ if and only if $i \leq \frac{y_t k - n + y_t + k}{n+1}$.

2. If $y_t \geq k$ and $\frac{y_t k}{2n} \geq 1$ then using the part (1) of the lemma

$$P\left(\frac{y_t k}{2n} \leq a_t \leq \frac{y_t k}{n}\right) \geq P\left(0 \leq a_t < \frac{y_t k}{2n}\right) \quad (7)$$

and then if $k \leq \frac{y_t}{2}$, it follows that

$$\begin{aligned} \frac{1}{2} &= P\left(0 \leq a_t < \frac{y_t k}{2n}\right) + P\left(\frac{y_t k}{2n} \leq a_t \leq k\right) \\ &\leq 2P\left(\frac{y_t k}{2n} \leq a_t \leq k\right) \end{aligned}$$

whereas if $k > \frac{y_t}{2}$ then

$$\begin{aligned} \frac{1}{2} &= P\left(0 \leq a_t < \frac{y_t k}{2n}\right) + P\left(\frac{y_t k}{2n} \leq a_t \leq \frac{y_t}{2}\right) + P\left(\frac{y_t}{2} < a_t \leq k\right) \\ &\leq 4P\left(\frac{y_t k}{2n} \leq a_t \leq \frac{y_t}{2}\right) \end{aligned}$$

where the last inequality is true because of relation (7) and the fact that the interval $(\frac{y_t}{2}, k]$ contains at most twice as many integers as $[\frac{y_t k}{2n}, \frac{y_t}{2}]$ does.

3. Now if $y_t \geq k$ and $1 \geq \frac{y_t k}{n} < 2$ then part (1) of the Lemma says that $P(a_t = 0) \leq P(a_t = 1) \leq P(1 \leq a_t \leq \min\{k, \frac{y_t}{2}\})$ and therefore imitating the proof of part (2) one gets that $\frac{1}{2} \leq 3P(1 \leq a_t \leq \min\{\frac{y_t}{2}, k\})$.
4. In this case, it suffices to bound $P(a_t = 0)$.

$$\begin{aligned} P(a_t = 0) &= \frac{(n-k)(n-k-1)\dots(n-k-y_t+1)}{2n(n-1)\dots(n-y-t+1)} \leq \frac{1}{2} \left(1 - \frac{k}{n}\right)^{y_t} \\ &\leq \frac{1}{2e^{-\frac{y_t k}{n}}} \leq \frac{1}{2\sqrt{2}} \end{aligned}$$

where the last inequality holds because $\frac{\log 2}{2} \leq \frac{y_t k}{n}$. Therefore,

$$P\left(1 \leq a_t \leq \min\left\{k, \frac{y_t}{2}\right\}\right) \geq \frac{\left(\frac{1}{2} - \frac{1}{2\sqrt{2}}\right)}{2} = \frac{\sqrt{2}-1}{4\sqrt{2}}$$

5. As above $P\left(\min\{\frac{y_t}{2}, k\} < a_t \leq k\right) \leq P\left(1 \leq a_t \leq \min\{\frac{y_t}{2}, k\}\right)$ and the goal is to bound $P(a_t = 0)$. Using the fact that $e^{-2x} \leq 1 - x$ whenever $x \leq \frac{\log 2}{2}$, conclude that

$$P(a_t = 0) \leq \frac{1}{2}e^{-2\frac{y_t k}{2n}} \leq \frac{1}{2}\left(1 - \frac{y_t k}{2n}\right)$$

and therefore

$$P\left(1 \leq a_t \leq \min\{\frac{y_t}{2}, k\}\right) \geq \frac{y_t k}{8n}$$

6. The proof of this case is similar to the proof of part (2) of the Lemma. The only difference is that a_t runs between 0 and y_t .
7. In this case $P(a_t = 0) \leq P(a_t = 1) \leq P\left(1 \leq a_t \leq \frac{y_t}{2}\right)$ because of part (1) of the Lemma. Then notice again that

$$P\left(\frac{y_t}{2} < a_t \leq y\right) \leq 2P\left(1 \leq a_t \leq \frac{y_t}{2}\right)$$

and then imitate the proof of part (2) of the Lemma.

8. Similarly to the cases above we have that $P\left(1 \leq a_t \leq \frac{y}{2}\right) \geq P\left(\frac{y}{2} \leq a_t \leq y\right)$. To bound $P(a_t = 0)$, expand as :

$$\begin{aligned} P(a_t = 0) &= \frac{(n-k)(n-k-1)\dots(n-k-y_t+1)}{2n(n-1)\dots(n-y-t+1)} \leq \frac{1}{2}\left(1 - \frac{k}{n}\right)^{y_t} \\ &\leq \frac{1}{2}e^{-\frac{y_t k}{n}} \leq \frac{1}{2\sqrt{2}} \end{aligned}$$

where the last inequality holds because $\frac{\log 2}{2} \leq \frac{y_t k}{n}$. Using the above and imitating the arguments from the above parts we have that $P\left(1 \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{(1-\frac{1}{\sqrt{2}})}{4} = \frac{\sqrt{2}-1}{4\sqrt{2}}$

9. As above $P\left(1 \leq a_t \leq \frac{y}{2}\right) \geq P\left(\frac{y}{2} \leq a_t \leq y\right)$. To bound $P(a_t = 0)$ use the fact that $e^{-2x} \leq 1 - x$ whenever $x \leq \frac{\log 2}{2}$. Then, the calculations of the previous part of the Lemma give that

$$P(a_t = 0) \leq \frac{1}{2}e^{-2\frac{y_t k}{2n}} \leq \frac{1}{2}\left(1 - \frac{y_t k}{2n}\right)$$

and therefore

$$P\left(1 \leq a_t \leq \frac{y_t}{2}\right) \geq \frac{y_t k}{8n}$$

□

The above lemma now leads to the proof of Theorem 1:

Proof. At first, in case that the starting number of the mismatched coordinates of the chains is odd wait until the coupling suggests staying fixed at one of them and taking a step on the other to turn the difference even. Call T_w the time the above happens. Then T_w follows a geometric distribution with probability of success $\frac{1}{2}$. So then $E(T_w) = 2$ and $Var(T_w) = 2$.

For general k , let y_t be the mismatched coordinates at time t . Also let a_t be the number of mismatched coordinates picked at time $t + 1$ in running the coupling process. Consider the following cases:

1. If at time t it is true that $\frac{y_t k}{n} \geq 2$ then $P\left(\frac{y_t k}{2n} \leq a_t \leq \min\{\frac{y_t}{2}, k\}\right) \geq \frac{1}{8}$. Then consider a much slower process where whenever picking $\frac{y_t k}{2n} \leq a_t \leq \min\{\frac{y_t}{2}, k\}$ mismatched coordinates we act as if only $\lfloor \frac{y_t k}{2n} \rfloor$ were picked. Then the expected number of steps to either exhaust all of the mismatched coordinates or fall into one of the other cases will be at most $\frac{8n}{k} \log n$.
2. If $1 \leq \frac{y_t k}{n} < 2$ then $P\left(1 \leq a_t \leq \min\{\frac{y_t}{2}, k\}\right) \geq \frac{1}{6}$. Therefore working as before, the expected number of steps to either exhaust all of the mismatched coordinates or fall in one of the other cases will be at most $\frac{3n}{k}$.
3. If $\frac{y_t k}{n} < 1$ then $P\left(1 \leq a_t \leq \min\{\frac{y_t}{2}, k\}\right) \geq \frac{\sqrt{2}-1}{4\sqrt{2}}$ so again the expected number of steps to either exhaust all of the mismatched coordinates or fall in one of the other cases will be at most $\frac{2\sqrt{2}n}{(\sqrt{2}-1)k}$.

Putting the bounds together:

$$E(T) \leq \frac{8n}{k} \log n + \frac{3n}{k} + \frac{2\sqrt{2}n}{(\sqrt{2}-1)k} + 2$$

and

$$Var(T) \leq A \frac{n}{k} \log n$$

where A is a constant. Then Chebychev's Inequality yields the rest.

□

4.3 Lower Bound

The lower bound will be proved using the eigenvectors and eigenvalues for this Markov Chain. Theorem 6 of [4] (page 49) says that the eigenvalues are the Krawtchouck polynomials and the eigenvectors are the normalized Krawtchouck polynomials. To see this, notice that the irreducible representations of $(\mathbb{Z}/2\mathbb{Z})^n$ are indexed by vectors $\mathbf{a} \in (\mathbb{Z}/2\mathbb{Z})^n$ so that

$$\rho_{\mathbf{a}}(\mathbf{v}) = (-1)^{\mathbf{a} \cdot \mathbf{v}}$$

Therefore, the Fourier transform of P at $\rho_{\mathbf{a}}$ is

$$\widehat{P}(\rho_{\mathbf{a}}) := \sum_{\mathbf{v}} \rho_{\mathbf{a}}(\mathbf{v}) P(\mathbf{v}) = \frac{1}{2} + \frac{1}{2} \sum_{b=0}^k (-1)^b \frac{\binom{j}{b} \binom{n-j}{k-b}}{\binom{n}{k}}$$

where j denotes the number of coordinates of \mathbf{a} that are equal to one. According to Theorem 6 of [4], the eigenvalues of the transition matrix are exactly the $\widehat{P}(\rho_{\mathbf{a}})$, $\mathbf{a} \in (\mathbb{Z}/2\mathbb{Z})^n$. The corresponding (non-normalized) eigenfunction is $f_{\mathbf{a}}(x) = (-1)^{\mathbf{x} \cdot \mathbf{a}}$. Notice that all $\mathbf{a} \in (\mathbb{Z}/2\mathbb{Z})^n$ that have the same number of zeros give the same eigenvalue. Thus, if $|\mathbf{x}|$ denotes the number of ones of \mathbf{x} , the j^{th} Krawtchouck polynomials

$$f_j(\mathbf{x}) = \sum_{b=0}^{|\mathbf{x}|} (-1)^b \frac{\binom{|\mathbf{x}|}{b} \binom{n-|\mathbf{x}|}{j-b}}{\binom{n}{j}}$$

are eigenfunctions and their normalized form will be used to compute the lower bound for the mixing time.

Remember that the definition of the total variation distance is

$$||P - Q|| = \sup_A |P(A) - Q(A)|.$$

A specific set A will provide a lower bound. To find this lower bound, consider the normalized Krawtchouck polynomial of degree one $f(x) = \sqrt{n}(1 - \frac{2x}{n})$ and the non-normalized Krawtchouck polynomial of degree two $f_2(x) = 1 - \frac{4x}{n-1} + \frac{4x^2}{n^2-n}$. Then, consider $A_{\alpha} = \{x : |f(x)| \leq \alpha\}$. A specific choice for α will guarantee the correct lower bound.

The orthogonality relations that the normalized Krawtchouck polynomials satisfy give that if Z is a point chosen uniformly in $X = \{0, 1, 2, \dots, n\}$ then

$$E_U\{f(Z)\} = 0 \text{ and } Var_U\{f(Z)\} = 1.$$

Now under the convolution measure,

$$E\{f(Z_l)\} = \sqrt{n} \left(\frac{1}{2} + \frac{1}{2} \left(1 - \frac{2k}{n} \right) \right)^l = \sqrt{n} \left(1 - \frac{k}{n} \right)^l$$

because f is an eigenfunction of the Markov Chain corresponding to the eigenvalue $1 - \frac{k}{n}$. Again under the convolution measure,

$$\begin{aligned} \text{Var}\{f(Z_l)\} &= \\ \frac{n}{n} + \frac{n(n-1)}{n} \left(1 + \frac{2k^2 - 2kn}{n^2 - n} \right)^l - n \left(1 - \frac{k}{n} \right)^{2l} &= \\ 1 + (n-1) \left(1 - \frac{2kn - 2k^2}{n^2 - n} \right)^l - n \left(1 - \frac{k}{n} \right)^{2l} \end{aligned}$$

Recall that the first three (non-normalized) eigenfunctions of this Markov Chain are:

$$f_0(x) = 1, f_1(x) = 1 - \frac{2x}{n} \text{ and } f_2(x) = 1 - \frac{4x}{n-1} + \frac{4x^2}{n^2 - n}.$$

By direct computation $f_1^2(x) = \frac{1}{n}f_0(x) + \frac{n-1}{n}f_2(x)$. Combining this and the fact that f_2 corresponds to the eigenvalue $1 + \frac{2k^2 - 2kn}{n^2 - n}$ gives the claimed variance.

Now, take l of the form $\frac{n}{2k} \log n - c \frac{n}{k}$ (where $c > 0$),

1. First case to be considered is $k = \lfloor \frac{n}{d} \rfloor$ where d is a constant. To simplify the notation, say that $k = \frac{n}{d}$.

$$E\{f(Z_l)\} = \sqrt{n} \left(1 - \frac{k}{n} \right)^l = \sqrt{n} \left(1 - \frac{1}{d} \right)^l$$

and

$$\begin{aligned} \text{Var}\{f(Z_l)\} &= 1 + (n-1) \left(1 - \frac{2kn - 2k^2}{n^2 - n} \right)^l - n \left(1 - \frac{k}{n} \right)^{2l} \\ &= 1 + (n-1) \left(1 - \frac{2n(1 - \frac{1}{d})}{d(n-1)} \right)^l - n \left(1 - \frac{1}{d} \right)^{2l} \end{aligned}$$

$$\begin{aligned} &\leq 1 + (n-1) \left(1 - \frac{2(1-\frac{1}{d})}{d}\right)^l - n \left(1 - \frac{1}{d}\right)^{2l} = \\ &1 + (n-1) \left(\frac{d-1}{d}\right)^{2l} - n \left(\frac{d-1}{d}\right)^{2l} \end{aligned}$$

In this case if $l = 1/2 \log_{d/(d-1)}(n) - c$ then

$$E\{f(Z_l)\} \geq \left(\frac{d}{d-1}\right)^c$$

so if $0 < c < \frac{1}{4} \log_{d/(d-1)}(n)$ the expectation $E\{f(Z_l)\}$ can get big while

$$\text{Var}\{f(Z_l)\} \leq 2$$

2. If there is $0 < \epsilon < 1$ so that $k = O(n^\epsilon)$ then the mean becomes

$$E\{f(Z_l)\} = \exp\left(c + O\left(\frac{k \log n}{n}\right) + O\left(\frac{ck}{n}\right)\right)$$

which means that for $0 < c < \frac{1}{4} \log(n/k)$ this expectation is big. Similarly for the variance

$$\begin{aligned} \text{Var}\{f(Z_l)\} &= \\ &1 + (n-1) \exp\left(c \frac{2n-2k}{n-1} - \frac{n-k}{n-1} \log n + O\left(\frac{k}{n} \log n\right) + O\left(\frac{ck}{n}\right)\right) \\ &\quad - \exp\left(2c + O\left(\frac{k \log n}{n}\right) + O\left(\frac{ck}{n}\right)\right) \\ &\leq 1 + O\left(\frac{1}{n}\right) + e^{2c} \left(O\left(\frac{k \log n}{n}\right) + O\left(\frac{ck}{n}\right)\right). \end{aligned}$$

Therefore the variance is uniformly bounded for $0 \leq c < \frac{1}{4} \log(n/k)$.

In both cases, Chebyshev's inequality gives that for the set $A_\alpha = \{x : |f(x)| \leq \alpha\}$,

$$U(A_\alpha) \geq 1 - \frac{1}{\alpha^2} \text{ while } P^{*l}(A_\alpha) < \frac{B}{(e^{2c} - \alpha)^2}$$

where B is uniformly bounded when $0 \leq c \leq \frac{1}{4} \log n$.

Therefore,

$$\|P^{*l} - U\|_{T.V.} \geq 1 - \frac{1}{\alpha^2} - \frac{B}{(e^{2c} - \alpha)^2}.$$

Now, take $\alpha = \frac{e^{2c}}{2}$, which finishes the proof.

5 Fourier Transform Arguments

In this section, a different approach is introduced. It combines the representation theory of the hypercube and the Fourier Transform to provide a bound for the mixing time. All of the irreducible representations of the hypercube are one dimensional and they are indexed by $z \in (\mathbb{Z}/2\mathbb{Z})^n$ in the following way:

$$\rho_{\mathbf{z}}(\mathbf{w}) = (-1)^{\mathbf{z} \cdot \mathbf{w}}$$

where $\mathbf{z} \cdot \mathbf{w}$ is the inner product of $\mathbf{z}, \mathbf{w} \in (\mathbb{Z}/2\mathbb{Z})^n$.

The Fourier Transform of a probability P at a representation ρ is defined as:

$$\hat{P}(\rho) = \sum_{g \in (\mathbb{Z}/2\mathbb{Z})^n} P(g) \rho(g)$$

which in our case means

$$\hat{P}(\rho_{\mathbf{z}}) = p + (1-p) \sum_{a=0}^j (-1)^a \frac{\binom{j}{a} \binom{n-j}{k-a}}{\binom{n}{k}} = p + (1-p) K_j^n(k) \quad (8)$$

where j is the number of ones that \mathbf{z} has and $K_j^n(k)$ is the j^{th} Krawtchouk polynomial evaluated at k .

In Chapter 3 of [4], one can find the Upper Bound Lemma (Lemma 1 in the book) which shows how using the Fourier transform of the representations of a group to find an upper bound for the mixing time of a walk on the group. More precisely, the upper bound lemma in the case of the hypercube (or in general for $(\mathbb{Z}/p\mathbb{Z})^\times$ says:

Lemma 10. (*Upper Bound Lemma*) *For a random walk on the hypercube, after l steps:*

$$4\|P^{*l} - U\|_{T.V.}^2 \leq \sum_{z \neq 0} (\hat{P}(\rho_z))^{2l}$$

6 The case $k=n/2$

In the case where n is even with $n = 2k$ where k is a positive, odd integer, the following facts hold:

Lemma 11. *For $k = \frac{n}{2}$ the Fourier Transform of representation $\rho_{\text{textbf{z}}$ is given by*

$$\widehat{P}(\rho_{\mathbf{z}}) = \begin{cases} \frac{1}{2}, & \text{if } j \text{ is odd} \\ \frac{1}{2} + \frac{(-1)^i \binom{\frac{n}{2}}{i}}{2 \binom{n}{2i}}, & \text{if } j=2i \end{cases}$$

where j is the number of ones that \mathbf{z} has.

Proof. According to Koekoek and Swarttouw in [10] the j^{th} Krawtchouck satisfies the following recurrence relation:

$$-kK_j^n(k) = \frac{1}{2}(n-i)K_{j+1}^n(k) - \frac{n}{2}K_j^n(k) + \frac{n}{2}K_{j-1}^n(k)$$

which for $k = \frac{n}{2}$ gives that

$$K_j^n\left(\frac{n}{2}\right) = \begin{cases} 0, & \text{if } j \text{ is odd} \\ \frac{(-1)^i \binom{\frac{n}{2}}{i}}{\binom{n}{2i}}, & \text{if } j=2i \end{cases}$$

given that $K_0^n(k) = 1$ and $K_1^n(k) = 1$. □

The next step is to bound the eigenvalues and use the Upper Bound Lemma to actually get an upper bound for the L^2 norm:

Lemma 12. *For every representation $\rho_{\mathbf{z}}$ where $\mathbf{z} \neq \mathbf{0}$,*

$$|\widehat{P}(\rho_{\mathbf{z}})| \leq \frac{3}{4} \tag{9}$$

Proof. If j is the number of ones \mathbf{z} has then if j odd the theorem holds because $\widehat{P}(\rho_{\mathbf{z}}) = \frac{1}{2}$. If $j = 2i$ the quantity $\frac{1}{2} + \frac{(-1)^i \binom{\frac{n}{2}}{i}}{2 \binom{n}{2i}}$ is the main concern. For i odd, the second term is negative but bigger than $-1/2$ therefore the quantity is positive less than $\frac{1}{2}$. For i even it turns out that $c_i = \frac{\binom{\frac{n}{2}}{i}}{\binom{n}{2i}}$ is maximized for $i = \frac{n}{2} - 1$ and at most $\frac{1}{n-1}$ by a simple argument. Thus for i even, $\frac{1}{2} + \frac{(-1)^i \binom{\frac{n}{2}}{i}}{2 \binom{n}{2i}} \leq \frac{3}{4}$. □

Theorem 13. For $k = \frac{n}{2}$ and for $l = \frac{n \log 2 - \log \epsilon}{\log \frac{4}{3}}$ with $0 < \epsilon < 1$,

$$4\|P^{*l} - U\|_{T.V.}^2 < \epsilon$$

Proof. After having computed the Fourier transform of each representation and bounded them in Lemma 12, the upper bound lemma gives:

$$4\|P^{*l} - U\|_{T.V.}^2 \leq \sum_{j=1}^n \binom{n}{j} b_j^{2l} \leq 2^n \left(\frac{3}{4}\right)^{2l} \leq \epsilon$$

for $l = \frac{n \log 2 - \log \epsilon}{\log \frac{4}{3}}$. □

Remark 14. Notice that for the l^2 norm,

$$\|G\| \|P^{*l} - U\|_2^2 \geq \left(\frac{1}{2}\right)^{2l} \sum_{i \text{ is odd}} \binom{n}{i} = 2^{\frac{n-1}{2} - 2l}$$

so indeed the l^2 norm cannot give a better upper bound for the mixing time. However, the l^1 norm may still be small for smaller l .

7 Similar Random Walks

This section focuses on similar random walks on $(\mathbb{Z}/m\mathbb{Z})^n$. It uses Fourier Transforms and comparison theory methods to bound their mixing times.

8 A random walk on $(\mathbb{Z}/m\mathbb{Z})^n$

Consider the walk generated by the measure

$$Q(a_{i_1} e_{i_1} + a_{i_2} e_{i_2} + \dots + a_{i_k} e_{i_k}) = \frac{1}{\binom{n}{k} m^k}$$

where $a_{i_j} \in \mathbb{Z}/m\mathbb{Z}$ and $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$.

Here is a proof of Theorem 3:

Proof. Let ρ_a denote a representation of $(\mathbb{Z}/m\mathbb{Z})^n$, where $a \in (\mathbb{Z}/m\mathbb{Z})^n$. If $g = (g_1, \dots, g_n) \in (\mathbb{Z}/m\mathbb{Z})^n$ then

$$\rho_a(g) = e^{\frac{2\pi i \sum_{j=1}^n a_j g_j}{m}}$$

Then the Fourier transform of Q at this representation with respect to Q is

$$\widehat{Q}(\rho_a) = \frac{1}{\binom{n}{k} m^k} \sum_{|g| \leq k} e^{\frac{2\pi i \sum_{j=1}^n a_j g_j}{m}}$$

where $|g|$ denotes the number of positions that g_i is not equal to zero. Since

$$\sum_{b \in \mathbb{Z}/m\mathbb{Z}} e^{\frac{2\pi i a_j b}{m}} = m \delta_{0, a_j},$$

if $n - |a| \geq k$,

$$\widehat{Q}(\rho_a) = \frac{\binom{n-|a|}{k}}{\binom{n}{k}}$$

otherwise $\widehat{Q}(\rho_a) = 0$. These are the eigenvalues of the random walk that are not equal to 1.

Now notice that all of the eigenvalues are non-negative and in particular

$$\begin{aligned} \frac{\binom{n-|a|}{k}}{\binom{n}{k}} &= \left(1 - \frac{k}{n}\right) \left(1 - \frac{k}{n-1}\right) \dots \left(1 - \frac{k}{n-|a|+1}\right) \leq e^{-k \sum_{j=n-|a|+1}^n \frac{1}{j}} \\ &\leq e^{-k \log \frac{n+1}{n-|a|+1}} = \left(1 - \frac{|a|}{n+1}\right)^k \end{aligned}$$

Then, the Upper Bound Lemma (Lemma 10) gives that

$$\begin{aligned} 4\|Q^{*l} - U\|_{T.V.}^2 &\leq \sum_{j=1}^{n-k} \binom{n}{j} (m-1)^j \left(1 - \frac{j}{n+1}\right)^{2kl} \\ &\leq \sum_{j=1}^{n-k} \frac{n^j}{j!} (m-1)^j \left(1 - \frac{j}{n+1}\right)^{2kl} \leq e^{-c} \end{aligned}$$

if $l = \frac{n+1}{2k} \log(mn) + \frac{c(n+1)}{2k}$. □

Remark 15. Notice that T , the first time that all coordinates have been touched is a strong stationary time, which implies that the total variation distance needs order $\frac{n}{k} \log n$ steps to get small. To see this one can imitate

the calculation for the coupon collector problem as presented in Lemma 2 of [3]. Further, notice that

$$|G| \|Q^{*l} - U\|_2^2 \geq n(m-1) \left(1 - \frac{1}{n}\right)^{2kl}$$

which means that the l^2 norm needs at least $\frac{n}{2k} \log(mn) + \frac{c(n+1)}{2k}$ steps to get small. Therefore there is a gap between the separation distance and the l^2 norm mixing times.

9 Comparison Theory Application

Comparison theory can help provide an upper bound for the following example:

Example 16. With notation as in Theorem 3, consider the case $k = 1$. Then,

$$Q(be_i) = \frac{1}{mn}$$

for $b \in \mathbb{Z}/m\mathbb{Z}$ and $1 \leq i \leq n$. The walk is "pick a coordinate at random and randomize it". Theorem 3 states that if $l = \frac{1}{2}((n+1) \log(mn) + c(n+1))$ then

$$4 \|Q^{*l} - U\|_{T.V.}^2 \leq e^{-c}$$

But then comparison theory gives the following theorem for the mixing time of the random walk generated by

$$P_2(\pm e_1) = \frac{1}{4n}, P(id) = \frac{1}{2}$$

Theorem 17. Let $l = \frac{1}{2}m^2((n+1) \log(mn) + c(n+1))$ then

$$4 \|P_2^{*l} - U\|_{T.V.}^2 \leq (1 + \frac{1}{n^n}) e^{-c}$$

Proof. Let $S = \{\pm e_j, id\}$ and $S' = \{be_j, b \in \mathbb{Z}/m\mathbb{Z}\}$. According to P. Diaconis and L. Saloff-Coste [6] if we represent each $z \in S'$ as a product of elements of S that has odd length and

$$A = \max_{s \in S} \frac{1}{P_2(s)} \sum_{z \in S'} \|z\| N(z, s) Q(z)$$

where $\|z\|$ is the length of this representation and $N(z, s)$ is the number of times s is inside the representation of z then

$$4\|P_2^{*l} - U\|_{T.V.}^2 \leq m^n e^{-l/A} + m^n \|Q^{*/2A} - U\|_2^2$$

An easy argument shows that $A \leq \max\{m^2, \frac{2}{m} + 2m\} = m^2$ therefore if $l = \frac{m^2}{2}((n+1)\log(mn) + c(n+1))$,

$$4\|P_2^{*l} - U\|_{T.V.}^2 \leq \frac{e^{-c}}{n^n} + e^{-c}$$

□

10 Acknowledgements

I would like to thank Persi Diaconis and Graham White for all the comments and discussions.

References

- [1] David Aldous. Minimization algorithms and random walk on the d -cube. *Ann. Probab.*, 11(2):403–413, 1983.
- [2] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.
- [3] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5):333–348, 1986.
- [4] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.
- [5] Persi Diaconis, R. L. Graham, and J. A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms*, 1(1):51–72, 1990.
- [6] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.*, 21(4):2131–2156, 1993.

- [7] Persi Diaconis and Mehrdad Shahshahani. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.*, 18(1):208–218, 1987.
- [8] Paul Ehrenfest and Tatjana Ehrenfest. über zwei bekannte einwände gegen das boltzmannsche h-theorem. *Physikalische Zeitschrift*, 8:311–314, 1907.
- [9] Mark Kac. Random walk and the theory of Brownian motion. *Amer. Math. Monthly*, 54:369–391, 1947.
- [10] Roelof Koekoek and Rene Swarttouw. *The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue*.
- [11] Evita Nestoridi and Graham White. Shuffling large decks of cards. <http://arxiv.org/abs/1606.01437>, 2016.
- [12] Laurent Saloff-Coste. Random walks on finite groups. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 263–346. Springer, Berlin, 2004.
- [13] Laurent Saloff-Coste. Total variation lower bounds for finite Markov chains: Wilson’s lemma. In *Random walks and geometry*, pages 515–532. Walter de Gruyter GmbH & Co. KG, Berlin, 2004.